Portland State University

PDXScholar

Geography Faculty Publications and
Presentations

Geography

7-11-2015

Evaluating User Interaction with a Web-Based Group Decision Support System: A Comparison Between Two Clustering Methods

Martin Swobodzinski

Portland State University, swobod@pdx.edu

Piotr Jankowski

San Diego State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/geog_fac

Part of the [Spatial Science Commons](#)

Let us know how access to this document benefits you.

Citation Details

Swobodzinski, M., & Jankowski, P. (2015). Evaluating user interaction with a web-based group decision support system: A comparison between two clustering methods. *Decision Support Systems*, 77, 148–157.

This Post-Print is brought to you for free and open access. It has been accepted for inclusion in Geography Faculty Publications and Presentations by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Accepted Manuscript

Evaluating user interaction with a web-based group decision support system:
A comparison between two clustering methods

Martin Swobodzinski, Piotr Jankowski

PII: S0167-9236(15)00120-7
DOI: doi: [10.1016/j.dss.2015.07.001](https://doi.org/10.1016/j.dss.2015.07.001)
Reference: DECSUP 12626

To appear in: *Decision Support Systems*

Received date: 7 September 2013
Revised date: 24 June 2015
Accepted date: 1 July 2015

Please cite this article as: Martin Swobodzinski, Piotr Jankowski, Evaluating user interaction with a web-based group decision support system: A comparison between two clustering methods, *Decision Support Systems* (2015), doi: [10.1016/j.dss.2015.07.001](https://doi.org/10.1016/j.dss.2015.07.001)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Evaluating user interaction with a web-based group decision support system: A comparison between two clustering methods

Martin Swobodzinski^{a,*}, Piotr Jankowski^{b,c}

^aDepartment of Geography, Portland State University, Portland, OR, United States

^bDepartment of Geography, San Diego State University, San Diego, CA, United States

^cInstitute of Geoecology and Geoinformation, Adam Mickiewicz University, Poznań, Poland

Abstract

Task-Technology Fit theory and the Technology Acceptance Model identify system utilization as an important indicator for the performance of complex software systems. Yet, empirical evaluations of user interaction with group decision support systems are scarce and often methodologically underdeveloped. For this study we employed an exploratory evaluation of user interaction in the context of web-based group decision support systems. Specifically, we used information-rich server logs captured through a web-based platform for participatory transportation planning to identify groups of users with similar use patterns. The groups were derived through multiple sequence alignment and hierarchical cluster analysis based on varying user activity measures. Subsequently, we assessed the reliability of the classifications obtained from the two clustering methods. Our results indicate limited reliability of classifications of activity sequences through multiple sequence alignment analysis and robust groupings from hierarchical cluster analysis for user activity initiations and durations. The presented work contributes a novel methodological framework for the evaluation of complex software systems that extends beyond the common approach of soliciting user satisfaction.

Keywords: web-based group decision support systems, human-computer interaction, server log analysis, use pattern evaluation, cluster analysis

1. Introduction

Ideally, the evaluation of complex group decision support systems (GDSS) would be part of an iterative process that aims at improving such systems with close involvement of the users [1]. User-centered evaluations of GDSS, however, require considerable resources to be committed in order to provide insight related to the suitability of GDSS for a given task [2]. Furthermore, evaluations of GDSS based on solicited user satisfaction are complicated by the potential conflation of individual satisfaction related to the decision-making process, the decision-making outcomes, and the decision support technology [3, 4]. Specifically, one can stipulate that a non-favorable decision-making outcome would lead to lower satisfaction related to the decision support tools that were employed in the process.

Task-Technology Fit (TTF) theory [5, 6, 7, 8, 4] and the *Technology Acceptance Model* [9, 10, 11] state that perceived usefulness informs engagement with and utilization of decision support technologies [10, 5, 12]. As such, system utilization constitutes an important observation by which to evaluate the performance of a decision support system for a given user base and task. Clearly, decision support tools need to be homogenized with the decision-making process in order for users to perceive the tools as useful [13].

Whereas user satisfaction can be elicited through survey methods, empirical evaluations of GDSS based on user interaction often defy prescriptive approaches of analysis [14, 15]. In this paper, we report on an exploratory empirical analysis of user interaction with a web-based GDSS for participatory transportation planning. The interaction of individuals with the GDSS was captured in information-rich server log files. Based on the server log files, we derived groups of individuals

*Corresponding author. swobod@pdx.edu; Tel.: +1-503-725-3164; Fax: +1-503-725-3166

with similar user interaction using multiple sequence alignment and hierarchical cluster analysis. Subsequently, we assessed the reliability of the classifications from the two clustering approaches.

In the next section, we describe the project in which the data for the presented analysis was collected. In addition, we provide a description on sequence alignment analysis and hierarchical cluster analysis and give a brief overview of evaluation approaches for web-based GDSS. In section 3, we outline our methodology for the collection and analysis of human-computer interactions within the context of the web-based GDSS at hand. In section 4, we discuss the results of our analysis as well as its limitations. Finally, in section 5, we state our conclusions and give suggestions for future work.

2. Related work

2.1. Participatory Geographic Information System for Transportation

The web-based GDSS analyzed in this paper was designed and implemented as part of the Participatory Geographic Information System for Transportation (PGIST) project, a basic-science research project funded by the Division of Information and Intelligent Systems of the National Science Foundation (EIA-0325916). The objective of the PGIST project was the development of an internet platform for public participation in regional transportation improvement programming [16, 17]. The PGIST project culminated in the deployment of the *Let's Improve Transportation* (LIT) website within the *LIT Challenge*, a structured, 4-week-long participatory planning process on the web.

Guided by a small team of moderators, the LIT participants were tasked with deliberating, analyzing, and selecting transportation projects and funding mechanisms that would pay for the implementation of the projects. To facilitate this process, the LIT website provided a host of analytical and deliberative decision support tools. Contextualized in participatory transportation planning, the LIT website exemplifies a web-based, communications-driven GDSS [18].

2.2. Multiple sequence alignment and hierarchical cluster analysis

Multiple sequence alignment analysis originated in molecular biology as a technique for the computational analysis of DNA and protein sequences [19, 20], with non-computational approaches going as far back as the 1960s [21]. The primary purpose of sequence alignment analysis is the detection of conserved patterns that reflect evolutionary relationships among sets of sequences [22]. Its output is commonly an arrangement of sequences in a tree structure, with the leaves of the tree representing individual sequences.¹ The shorter the distance between individual leaves along the branches of the tree, the greater the computed similarity between sequences. The

topological structure of the tree can be used to derive clusters of similar sequences by grouping sequences that are located in the same vicinity of the tree.

Sequence alignment procedures are computationally demanding, with a running time in the order of $O(n^k)$, with k being the number of sequences and n their average length. As such, the simultaneous alignment of sequences can burden computational resources, even when processing a small number of sequences [23, 24, 25]. To mitigate such limitations, many multiple sequence alignment algorithms rely on an iterative approach that is based on a progression of pairwise alignments.

Similar to multiple sequence analysis, hierarchical cluster analysis also produces a classification tree. Rather than using the alignment of sequences as the basis for the calculation of a similarity measure, however, hierarchical cluster analysis derives similarity scores from numeric distances of ratio-level variables. As an agglomerative clustering algorithm, the cluster formation starts out with as many clusters as there are observations, with one observation per cluster. The algorithm iteratively computes distances between pairs of clusters and combines the clusters with the smallest distance (i.e., the most similar clusters). As to derive the final classification tree, various clustering strategies (e.g., nearest neighbor and average linkage between groups) and similarity measures (e.g., Euclidean distance) can be employed.

2.3. Sequence alignment analysis in the social sciences

Sequence alignment analysis found entry into the social sciences through the work of Abbott [27, 28, 26]. Specifically, researchers aimed at extending existing implementations of sequence alignment algorithms to allow for the coding of observations beyond DNA and proteins. Such efforts led to the creation of the sequence alignment analysis software ClustalG [29] and ClustalTX [30], both derivatives of Clustal, a prominent multiple sequence alignment algorithm in the bioinformatics community [31, 32, 33, 34, 35]. In addition to the work of Wilson on daily activity routines [36, 37] and activity-space trajectories [30], extended multiple sequence alignment algorithms have been applied to the scheduling of vacations [38], sightseeing behavior [39], and eye movement patterns in the context of static small multiple map displays [40].

2.4. Evaluation of web-based group decision support systems

Among a few examples of empirical evaluations of web-based GDSS, Chen et al. [41] reported on TeamSpirit, a web-based GDSS for problem solving by distributed teams. TeamSpirit was designed to support the Creative Problem Solving (CPS) process. The CPS process entails the exploration and definition of problems, the generation and evaluation of alternatives, and the planning for and evaluation of solutions. The evaluation of TeamSpirit focused on its use by different subjects in relation to the subjects' performance on varying problem-solving tasks, the amount of CPS training, and the amount of TeamSpirit training. Central to the evaluation of TeamSpirit were not individual human-computer interaction patterns, as is the case in our approach, but the effect of the usage of TeamSpirit on tasks external to the system. The results by Chen et al.

¹In the following, we will use the term sequence and leaf interchangeably.

[41] showed that the quality and quantity of generated ideas increased significantly for groups that received training.

In the context of the PGIST project, our colleagues reported on a spatiotemporal evaluation of discussion contributions within the LIT Challenge [42, 43]. In particular, the authors analyzed a subset of deliberation-related data on the creation of discussion posts and the expression of agreement and disagreement with such posts. Focusing on the visual analysis of online discussion in the LIT Challenge, the authors developed a visualization tool which overlaid a map of the study area with a graphical plot of the frequency of discussion contributions during the LIT Challenge. The visualization tool showed the physical locations of contributors as well as categories of responses. Analytical activities were only considered as temporal reference points for deliberation-related activities. The evaluation that we present in this paper, on the other hand, forgoes a particular emphasis on deliberation-related activities in favor of a holistic evaluation of all activities (i.e., deliberation, analysis, and information retrieval) that were captured within the LIT Challenge. Given the lack of prescriptive approaches to the empirical analysis of web-based GDSS, we embarked on an exploratory analysis of the data generated within the LIT Challenge. We opted for multiple sequence alignment analysis and hierarchical cluster analysis as these techniques were most suitable for the analysis of the LIT data.

3. Methods

3.1. Participants and collection of server logs

The pool of participants in the LIT Challenge was comprised of 246 individuals that were recruited proportionally from King, Pierce, and Snohomish County in the central Puget Sound region of the State of Washington. Of these, 179 were eligible for a small stipend as compensation for their participation. The amount of the stipend was cumulative and depended on the number of completed steps in the LIT Challenge, as well as the participation in data-gathering activities that were part of the research design (e.g., online questionnaires and user interviews). A subset of 47 individuals completed all steps in the LIT Challenge, that is, they interacted with LIT from the beginning to the end of the participatory planning process. These 47 individuals constitute the sample for the evaluation of the LIT GDSS that we present in this paper.

The participatory process was structured into five main steps which were divided into one to three sub-steps. The progression through the steps was subject to a set time schedule (Table 1). Within the constraints of the time schedule, the participants were interacting with the LIT website through a web browser. The LIT website offered a variety of tools to support the participants in their decision making about transportation projects. A detailed description of the conceptual framework and design of the LIT website, the structure of the LIT Challenge, as well as the recruitment of participants can be found in Young et al. [44], Lowry et al. [17], [16], and Nyerges and

Table 1: Overview of the LIT steps with their respective availability dates.

Description	Date
Step 1: Discuss your concerns	
1a Map your daily travel	10/16–10/18
1b Brainstorm concerns	10/16–10/18
1c Review summaries	10/19–10/22
Step 2: Assess improvement factors	
2a Review factors	10/23–10/25
2b Weigh factors	10/23–10/25
Step 3: Create transportation packages	
3a Discuss projects ^a	10/26–10/31
3b Discuss funding options ^a	10/26–10/31
3c Create your own package	10/29–10/31
Step 4: Select a package for recommendation	
4a Discuss candidate packages	11/01–11/06
4b Vote on candidate packages	11/05–11/06
Step 5: Prepare group report	
5a Review draft report	11/07–11/13
5b Vote on final report	11/11–11/13

^aavailable at any given time

Aguirre [43].²

In order to be able to partake in the LIT Challenge, participants had to create a user account with a unique username and password. By browsing the LIT website, the participants invoked associated functionality on the LIT server which was tracked and captured by server-side scripting procedures in a server log file. Each stored server log entry included a unique identifier for the user who invoked the functionality and a time stamp that captured the date and time of the invocation. In addition, data was collected that allowed us to determine the context in which the invocation occurred (e.g., the name of the invoked method and the submitted parameter values).

3.1.1. Server logs and user activities

The server captured HTTP messages related to invocations of objects, methods, and scripts that were accessible to the user through the web interface. A server log entry that was captured during the LIT Challenge is shown in Fig. 1. In many cases, a single invocation of functionality resulted in multiple entries in the server log file.

We found that this level of user interaction did not immediately lend itself to the analysis of user behavior since the logs were closely tied to the programmatic structure of the system. As a result, we devised algorithms that aggregated the server logs into user activities. A prerequisite for such an aggregation was the identification of distinguishable user activities. Distinguishable in this context refers to being able to determine the types as well as the beginning and end times of activities.

In most cases, the beginning of an activity was readily extractable from a particular server log entry. The end of an activity, on the other hand, required a more sophisticated handling of the server log entries. In many cases, the end of a given activity

²As of July 2015, the LIT website (www.letsimprovetransportation.org) is accessible through a public guest account to individuals interested in exploring the website.

Figure 1: An example of a server log entry (excluding non-crucial parameters).

<u>Timestamp</u>	<u>User_ID</u>	<u>URL</u>	<u>Session_ID</u>	<u>Referrer</u>
10/14/07 6:51:26 PM	39	usercp.do	C5BC0B286A23F335E4	main.do
⏟ Date and time of invocation	⏟ Unique user identifier	⏟ Method that was invoked	⏟ Unique session identifier	⏟ Method that was invoked before

was not captured through a terminal server log entry but rather determined through the start of a new activity, a log-out event, or the beginning of a new HTTP session.

At the most granular level, we were able to differentiate 25 user activities, as listed in Table 2. Out of the 25 activities, 22 were non-instantaneous activities (i.e., associated with duration). These activities functioned as the building blocks for the construction of individual user interaction sequences. In particular, we devised two sets of sequences: 1. representing the succession of user activities and 2. representing both the succession and duration of activities. We opted for a 1-second resolution for the activity durations since the timestamps of the server log entries were resolved at the same resolution.

In addition, we aggregated thematically related activities which resulted in a separate set of sequences with nine distinct activities, eight of which carried activity duration information³. Finally, we calculated the overall time that each user spent interacting with the LIT website and derived a single numeric measure of user interaction. This multi-granular approach resulted in three distinct measures of user interaction to be analyzed.

3.2. Sequence alignment software and sequence generation

The sequence alignment software that we employed in our analysis was ClustalTXY [30]. ClustalTXY requires input data to be formatted in a sequence of characters, or *words*, with a maximal length of six characters. The set of possible characters is comprised of the English alphabet. For the storage of the sequences of individual user interaction with the LIT website, we opted for the FASTA format—the de-facto standard for sequential data. FASTA is a text file format in which individual sequences are composed of a header (demarcated by ">"), followed by a unique identifier for the sequence and a count of the number of words in the sequence. The lines following the header contain the sequence itself (Fig. 2).

We devised a coding schema that assigned a unique three-character word to each user interaction activity (Table 3) and generated a total of four sequences of user interaction: 1. the successions of activities at the 25-activity level, 2. the successions and durations of activities at the 25-activity level, 3. the successions of activities at the 9-activity level, and 4. the successions and durations of activities at the 9-activity level. From the sequences that included user activity durations, we tabulated the total length of the various activities by parsing occurrences

of each word in any given sequence. The summation of all activity durations provided us with the total length of interaction with the LIT website that each participant engaged in.

For the sequences that captured successions of activities alone, the count of words represented the overall number of activity initiations by an individual user. Unfortunately, such measure cannot be operationalized within sequence alignment analysis. Even though one could devise sequences that are composed of identical characters (e.g., AAAA for user 1 and AA for user 2), such sequences would always lead to a perfect alignment, with the clustering tree collapsing into a single node—a trivial result.

Within our hierarchical cluster analysis, we analyzed both the counts of activity initiations and the measures of activity durations based on two difference measures.⁴ For the counts of activity initiations, we employed phi-square, a distance measure for count data based on a normalized form of chi-square.⁵ For the measures of activity durations (i.e., a continuous, ratio-level variable), we used squared Euclidean distance⁶. Furthermore, to determine the level of stability of the classification tree, we examined the results from hierarchical cluster analysis across different clustering methods [45]. An overview of the sequence alignment and hierarchical cluster analysis of the user interaction datasets is given in Table 4 and 5.

4. Results

4.1. Hierarchical cluster analysis results

For the hierarchical cluster analysis, we employed three levels of activity granularity (i.e., 25-activity level, 9-activity level, and overall activity) and two different types of user interaction

⁴Difference measure refers to a method of calculating the dissimilarity between clusters.

⁵Phi-square is derived by normalizing chi-square by the square root of the combined frequencies, as defined by:

$$dist_{Ph2}(x, y) = \sqrt{\frac{\frac{\sum_i (x_i - E(x_i))^2}{E(x_i)} + \frac{\sum_i (y_i - E(y_i))^2}{E(y_i)}}{N}}$$

where x, y are data points, $E(x_i), E(y_i)$ are the expected frequencies of the variables of the respective data points and N denotes the combined observed frequencies.

⁶As defined by:

$$dist_E(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$$

where x, y are data points and x_i, y_i are values of the variables of the data points.

³The voting activity was the only activity out of the nine that was instantaneous and as such did not have a meaningful temporal length or duration.

Table 2: User interaction activities overview for the fine (i.e., 25 activities) and aggregated (i.e., 9 activities) levels of analysis.

Type	Fine-level Activities	Aggregated Activities
Deliberation	reading post: projects	reading post
	reading post: funding	
	reading post: concerns	
	reading post: summary	
	reading post: factors	
	reading post: packages	
	reading post: draft	
	deliberating package vote	deliberating package vote
	creating post: projects	creating posts
	creating post: funding	
Analysis	creating post: concerns	
	creating post: summary	
	creating post: factors	
	creating post: packages	
	creating post: draft	
Information	weighting factors	weighting factors
	creating package	creating package
	voting: post	various voting
	voting: package	
Information	voting: report	
	mapping routes	mapping routes
	getting info: project	getting project/package info
	getting info: packages	
Information	getting info: draft	getting other info
	getting info: report	

Figure 2: An example of a user interaction sequence file in the FASTA format for two users with the identifier 13 and 26. The sequence of user 13 is comprised of 54 and the sequence of user 26 of 41 three-character words.

```

>13 54
MdrRpaCpaCpaCpaCpaVoaRpaVoaRpaRpaRpaRpaRpaRpaRpaRpaRpaRpaWit
RpaWitRpaGipCpaRpaRpaRpaRpaRpaVoaRpaCypCypCypCypCypCypRpaGip
GipRpaRpaRpaRpaRpaDpvRpaVoaRpaGioVoaGioVoa
>26 41
MdrRpaRpaVoaRpaVoaRpaRpaVoaVoaRpaVoaRpaVoaRpaCypCypRpaRpaVoa
RpaRpaGipGipGipRpaGipVoaDpvRpaVoaRpaVoaRpaVoaRpaVoaRpaVoaRpa
Rpa

```

Table 3: User interaction activities coding schema at the level of (a) 25 and (b) 9 activities.

(a) fine-level user interaction			(b) aggregated user interaction		
Activity		Code	Activity		Code
reading post:	projects	Rpr	reading post		Rpa
reading post:	funding	Rpf	deliberating package vote		Dpv
reading post:	concerns	Rpc	creating post		Cpa
reading post:	summary	Rps	weighting factors		Wit
reading post:	factors	Rpt	creating package		Cyp
reading post:	packages	Rpp	various voting		Voa
reading post:	draft	Rpd	mapping routes		Mdr
deliberating package vote		Dpv	getting project/package info		Gip
creating post:	projects	Cpr	getting other info		Gio
creating post:	funding	Cpf			
creating post:	concerns	Cpc			
creating post:	summary	Cps			
creating post:	factors	Cpt			
creating post:	packages	Cpp			
creating post:	draft	Cpd			
weighting factors		Wit			
creating package		Cyp			
voting: post		Vud			
voting: package		Vop			
voting: report		Voo			
mapping routes		Mdr			
getting info: project		Gri			
getting info: packages		Gpi			
getting info: draft		Gdi			
getting info: report		Goi			

Table 4: Overview of hierarchical cluster (HCA) and multiple sequence alignment analyses (MSA) of the LIT data in relation to the granularity of activities. Individual analyses are denoted with an x in the cells. Appropriate pairwise comparisons of the classification results from HCA and MSA are represented through arrows linking the respective activity analysis levels.

#Activities	Activity analysis levels			
	Hierarchical clustering		Sequence alignment	
	Initiation	Duration	Succession	Duration
25	x	x	x	x
9	x	x	x	x
overall	x	x	n/a	n/a

Table 5: User activity analyses overview showing the four possible combinations of the sequence alignment analysis in relation to the matched comparative hierarchical cluster analyses. The variable under analysis, the distance measure, and the clustering algorithms that were employed in the hierarchical cluster analyses are stated in the cells of the table.

Sequence alignment	Hierarchical cluster analysis		
	Variable	Difference measure	Clustering method
Succession	Activity initiations	Phi-square	between-group, within-group, nearest/furthest neighbor, median, centroid, Ward's
Duration	Activity durations	(Euclidean distance) ²	

data (i.e., frequency counts and activity durations in seconds). An icicle plot for the furthest neighbor clustering method is shown in Fig. 3.⁷ Along the x-axis, individual cases (i.e., data points/users) are identified by a number expressing the order in which an individual case was processed, followed by a unique identifier for that case (e.g., 28 User47). Each case is represented by a labeled, vertical column which is flanked by non-labeled columns that function as separators. The numbers of clusters within a specific cluster solution are captured along the y-axis.

In Fig. 3, possible cluster solutions range from 8 clusters at the bottom to 1 cluster along the upper edge of the icicle plot. For each integer mark along the y-axis, a horizontal line can be drawn from left to right that dissects the case and separator columns. For a given number of clusters, this cross-section allows us to determine the cases that are included in particular clusters. For instance, for mark 3 on the y-axis (i.e., a 3-cluster solution), users 47, 46, 42, and 44 comprise the first cluster and users 30, 38, 35, 36, 40, 31, and 27 the second cluster. The third cluster spans from user 1 to user 19. Similarly, at mark 2, the 2-cluster solution is comprised of a cluster that contains users 47, 46, 42, and 44, as in the 3-cluster solution, and a second cluster that includes the remaining 43 cases, from user 30 to user 19. A complete inventory of the icicle plots from the hierarchical cluster analysis is provided in the appendix.

4.1.1. Hierarchical cluster analysis results: activity initiations

For our sample of 47 cases, a classification with four clusters of similar size would include 10 to 11 cases per cluster. Such configuration would be desirable for subsequent statistical evaluations of the clusters (e.g., regression analysis of cluster membership and individual socio-demographic characteristics). In our assessment of the icicle plots, we paid particular attention to the overall appearance of the classification as well as the number of cases in clusters around the 4-cluster mark.

For the levels of 25 and 9 activities, we found similar patterns in the icicle plots. For a 4-cluster solution, the classification of the between-group, nearest neighbor, centroid, and median method generally consisted of a large single cluster that was accompanied by three single, non-classified cases (i.e., individual cases that were not incorporated into a larger cluster). There were minor deviations within the between-group and median icicle plots at the 9-activity level in which 2-case clusters appeared rather than single-case clusters. In all of these classifications, for a 4-cluster solution, a single cluster that included at least 40 of the 47 cases (i.e., 85%) was dominating the classification.

The icicle plots for the within-group, furthest neighbor, and Ward's method exhibited more evenly sized clusters and were generally less skewed than the icicle plots for the four other classification methods. The results from the within-group and the furthest neighbor method for 25 and 9 activities showed similar patterns, with their respective classifications consisting

of 3 to 4 major clusters that encompassed 8 to 24 cases. For both activity levels, Ward's method lead to 4 major clusters consisting of 5 to 15 cases.

In the icicle plot pertaining to the overall count of activity initiations, the classification patterns differed considerably from the 9 and 25-activity levels in that only the icicle plot for the nearest neighbor method resembled the plots seen with the between-group, nearest-neighbor, centroid, and median methods for 25 and 9 activities (i.e., a single large cluster accompanied by three non-classified cases). The nearest neighbor method also provided the weakest classification, that is, a classification with little discrimination between clusters. The icicle plots of the remaining clustering methods did not include single-case clusters, with cluster sizes ranging from 4 to 27 cases for a 4-cluster solution. The dominating pattern was a major single cluster that was accompanied by 4 minor clusters containing 3 to 10 cases.

4.1.2. Hierarchical cluster analysis results: activity durations

Compared to the rather diverse classifications resulting from the counts of activity initiations, we found less diversity for the measures of activity durations at the levels of 25 and 9 activities. With the exception of Ward's method for the 9-activity level, none of the other clustering methods resulted in a classification with sizable multi-case clusters for a 4-cluster solution. Instead, the classifications were dominated by a single large cluster containing at least 32 (68%) of the cases. The remaining cases were predominately assigned to small clusters with 3 cases or less, leading to a skewed appearance of the icicle plots. This pattern was more pronounced for the 25-activity level than for the 9-activity level.

At the overall activity level (i.e., the overall time spent interacting with the LIT website), the classifications featured a single-case cluster (comprised of user 47) and a three-case cluster (user 46, 42, and 44) across all clustering methods for a 4-cluster solution. For the within-group, nearest neighbor, and median method, a two-case cluster was formed by users 30 and 38. The latter cluster was subsumed in a larger cluster in other classifications. With the exception of Ward's method, a large cluster containing at least 36 (77%) of the cases dominated the classification. For Ward's method, we found four well-formed clusters with 4, 7, 15, and 21 cases.

4.1.3. Discussion of the hierarchical cluster analysis results

Our summary of the classification results for the hierarchical cluster analysis is given in Table 6. Compared to the other clustering methods, Ward's method⁸, resulted in well-differentiated clusters (i.e., strong classifications) for all activity levels, irrespective of whether the counts of activity initiations or the durations of activities were concerned. Such behavior is not surprising given that Ward's method is biased towards the creation of equally sized clusters. We see this bias as desirable for the subsequent analysis of the characteristics

⁷In order to conserve page space, we did not include an exhaustive set of icicle plots in this manuscript. These figures can be requested from the corresponding author instead.

⁸Also known as the minimum variance method [46].

Figure 3: An example of an icicle plot for the furthest neighbor clustering method. Unique user identifiers are shown along the x-axis and numbers of clusters along the y-axis. For instance, a two-cluster solution is comprised of (read from left to right) user 47, 46, 42, and 44 in the first and the remaining users in the second cluster.

Table 6: Hierarchical cluster analysis results for the three activity levels in relation to varying clustering methods. A white square indicates a weak classification (i.e., clusters of unbalanced size, including a number of single-case clusters) and a black square a strong classification (i.e., well-formed clusters of relatively comparable size). A black triangle represents a classification that is considered to be in-between weak and strong. A strong classification is desirable for a statistical analysis of the characteristics of the cases in the clusters.

Clustering method	Activity granularity					
	Activity initiations			Activity durations		
	25	9	1 ^a	25	9	1 ^a
between-groups	□	□	■	□	□	▲
within-groups	■	■	■	□	□	▲
nearest neighbor	□	□	□	□	□	▲
furthest neighbor	□	▲	■	□	▲	▲
centroid	□	□	▲	□	□	▲
median	□	□	■	□	▲	▲
Ward's	■	■	■	▲	■	■

^adenotes the overall level of user interaction

of the individuals that were clustered together. For the human-computer interaction of the LIT participants, the results from Ward's method conformed to our expectation of group-based differences in user interaction among the 47 individuals under analysis, especially at the overall activity level capturing duration of interaction. In particular, our results showed that the 47 individuals were classified into sizable, disparate clusters which suggests moderate similarity between groups of LIT participants in terms of overall duration of interaction.

In order to interpret the composition of the clusters, the clusters could be subjected to further analysis through multivariate statistical methods. The clusters, however, might bear uncertainty in regard to the reliability of the classification. Mitigating such concern for Ward's method is its well-documented strength of reliably identifying clusters [e.g., 47, 48]. Even though the classification results from Ward's method stood out as exceptional at the 25 and 9-activity level, at the overall activity level, a majority of the seven clustering methods mimicked the classification patterns obtained from Ward's method for both activity initiations and activity durations. Thus, we concluded that Ward's method for the overall activity level indicated the existence of reliable multi-case clusters for a 4-cluster solution.

4.1.4. Limitations of the hierarchical cluster analysis

Caution should be taken when interpreting the results from the phi-square-based analysis of differences between counts of activity initiations. In particular, given the small size of our

dataset, there was a sizeable number of cells within the table of counts of activity initiations that did not conform to the minimal count requirement for the chi-square distribution. The minimal count requirement specifies that most cells should have a value greater than 5 and no cell should contain a value less than 3. The non-conformity was most evident for the level of 25 activities and slightly less in the case of 9 activities. In the case of 25 activities, 78% of the cells had a value of less than 5; for 9 activities, this percentage dropped to 63%. At the overall activity level, there were no such cells. However, since the overall pattern of a stronger classification for the overall activity level repeated itself in both the phi-square and Euclidean distance based clustering, we posit that greater data variability was decisive for the classification rather than the sparseness of the data.

4.2. Sequence alignment analysis results

For the multiple sequence alignment analysis, we employed two levels of activity granularity (i.e., 25 and 9 activities), and two different types of individual-level user interaction data: sequences coding the succession of activities for each individual and sequences capturing both succession and duration of activities. In the following, we present the results from the analysis of these two datasets.

4.2.1. Sequence alignment analysis results: activity successions

In order to assess the stability of the classification that was output by the multiple sequence alignment software, we employed a bootstrap approach. The bootstrap algorithm randomly samples components in the set of aligned sequences so that the size of the samples corresponds to the length of the sequences that were observed. Next, a new tree is derived from the samples which is then compared to the classification tree. This procedure is repeated many times; in our case, we opted for the default value of 1000 repetitions. Subsequently, the topology of the bootstrap trees is compared to the classification tree. Corresponding bifurcations in the trees are counted and the overall sum of correspondence (i.e., the bootstrap value) is assigned as a value to the respective bifurcations of the classification tree. A bifurcation is considered stable if it reproduces in 95% of all bootstrap trees [49].

The results for the analysis of the sequences of activity successions for the two activity levels are shown in Fig. 4. For 25 activities (Fig. 4a), we found the majority of bootstrap values to be considerably less than the threshold value of 950 (i.e., 95% of 1000 bootstrap iterations). As expected, the stability of the bifurcations increased considerably towards the leaves, with the bootstrap values ranging from 308 to 990 for the bifurcations before the leaf-level. Only for three bifurcations, all preceding the leaf-level, did the bootstrap value exceed 950. Similarly, for 9 activities (Fig. 4b), the terminal bifurcations carried the largest bootstrap values, ranging from 111 to 998. For both activity levels, we found very small bootstrap value towards the center of the tree (i.e., the leftmost part of the tree)⁹. Around

the center of the tree, the distances (i.e., dissimilarity) between individual branches were generally the smallest, ranging from 1 to 479 for the first three bifurcations from the center. The number of cases that were clustered in the branches spawning from these bifurcation were the largest.

4.2.2. Sequence alignment analysis results: activity durations

The results of the analysis of sequences that captured activity durations are shown in Fig. 5. In general, the weak stability of higher-level branches (i.e., those close to the center of the tree) that we found for activity successions was not as prominent for activity durations. For the 9-activity level, bifurcations that surpassed the threshold value of 950 consisted mostly of 2-case clusters. Bifurcations towards the center of the tree generally had bootstrap values of 278 or less.

We observed a remarkable exception at the level of 25 activities (Fig. 5a). In regard to the bootstrap values, the classification suggested very high stability across large parts of the tree, with 35 out of the 44 bifurcations (80%) having a bootstrap value above 950. Out of these 35 bifurcations, 29 carried a bootstrap value of 1000.

4.2.3. Discussion of the sequence alignment analysis results

The bootstrap results showed stark differences between the 25 and 9-activity levels. For two sets of sequences in which the overall information content (i.e., the length of sequences) was kept constant, we expected the amount of variability to decrease in the sequences of lower activity granularity¹⁰. In turn, we considered it likely that the stability of the classification tree were to increase from the 25-activity to the 9-activity level. Similarly, we expected to see generally lower bootstrap value for the classification tree for activity durations compared to activity initiations due to the sequences being much longer¹¹. This amplified the differences in lengths among individual sequences which can negatively impact the quality of the alignment [50]. However, contrary to our expectations, we observed a considerable increase in the stability of the classification tree as indicated by the large bootstrap values at the 25-activity level for activity durations. For that level, the sizable number of distant, multi-case clusters in the classification tree in combination with the significant bootstrap values (i.e., greater than 950) suggested pronounced dissimilarity in user interaction patterns among small groups of LIT participants.

The particularly high bootstrap values at the 25-activity level for activity durations were concerning. Apart from the visual assessment of the trees (and possibly a follow-up analysis of the topology of the individual trees), the bootstrap mechanism acts as an important indicator for the reliability of the clustering results obtained from multiple sequence alignment. As stated before, one would expect to see a weaker stability for activity durations than activity successions since the latter contains less variability. In addition, the sequences that were aligned in the

⁹The classification trees in multiple sequence alignments analysis are unrooted.

¹⁰Here: the 9-activity level

¹¹The longest sequence for activity durations contained 41000 words whereas the longest for activity successions contained 1000 words.

Figure 4: Bootstrap results for the activity successions (i.e., no duration) of 25 and 9 activities. The leaves of the tree denote unique case (i.e., user) IDs. The lengths of the tree branches convey the standardized distance (i.e., dissimilarity) between cases, with values ranging from 0 to 0.5, in increments of 0.1.

activity succession case were much shorter than those in the activity durations case.

These results from the multiple sequence alignment analysis also ran contrary to the results obtained from the hierarchical cluster analysis. In the latter, none of the datasets at the 25-activity level provided for a strong classification, with the 25-activity level of activity durations in fact resulting in the weakest classification among all datasets. Certainly, it is possible that the sequence alignment algorithm could have provided an exceptionally strong classification for the most variable dataset. In such a case, however, we would have expected to find comparably high bootstrap values for both the 25-activity and 9-activity levels rather than the 25-activity level alone. Instead, the bootstrap values for the 9-activity level did not exceed the threshold of 950 for clusters with more than 2 cases. Hence, it was questionable that sequence alignment analysis could be relied upon as far as identifying LIT users with similar user interaction patterns is concerned. We recommend that investigations be conducted to assess the performance of sequence alignment analysis for well-defined sequences of behavioral user interaction data, especially for sets of sequences where lengths vary greatly. As a viable alternative to sequence alignment analysis for scenarios similar to ours, we suggest the use of hierarchical cluster analysis.

4.2.4. Limitations of the sequence alignment analysis

The 47 participants that we analyzed were eligible for a small monetary compensation. We do not have data, however, related to the effect of the stipend on the behavior of these participants. We believe that the prospect of monetary gain played a minor role at most, considering that the LIT Challenge afforded a commitment for more than a month. In addition, such incentive was shared among the 47 participants which, so to speak, should have controlled for monetary compensation on behavior, at least in this particular group. That is not to say that there might not have been a cluster of individuals whose main motivation was to get some monetary reward for participating in the LIT Challenge. Such insight, however, would be based on an interpretation of the clusters that we suggest would be worthwhile to undertake following the approach described in this paper.

The weight matrix that we employed within the sequence alignment analysis to derive similarity scores between sequences was an identity matrix. In other words, there was no representation of qualitative similarities between activities as all activities were considered separate and unique. Intuitively, various activities (e.g., reading comments related to individual concerns and reading comments in regard to transportation projects) that we distinguished within the server logs were presumably similar in terms of the motivation that individual subjects might have had for engaging in the activity. To make such a similarity workable within the sequence alignment analysis would require the quantification and numerical representation of similarities among activities within the weight matrix. This raises questions in regard to epistemology, semantics, and representation of human behavior which are beyond the scope of this paper. At the least, numerical representation, if feasible, would be

contingent upon an in-depth understanding of the factors that drive individual choice making in the context of participatory planning—which is a grand challenge that necessitates collaboration among many different research communities.

It is unclear if the weak classifications for the 25 and 9-activity levels resulted from genuine similarity of the user interaction patterns (possibly due to the nature of the participatory decision-making process and the LIT user interface). Given the strong classifications that we obtained for the overall activity level, changes to the temporal resolution and/or granularity of activities might provide further insight into the behavioral user interaction patterns of the LIT participants. For instance, sequences could be generated that capture user interaction for a certain time frame rather than the whole duration of the LIT Challenge. Similarly, changes to the aggregation of activities (e.g., activities classified into analytical and deliberative activities) would provide different numeric measures of user interaction that could be subjected to hierarchical cluster analysis. Alternatively, differences in user interaction could be assessed based on comparisons of clustering results that focus on a subset of activities (e.g., all reading-related activities) as opposed to the exhaustive set of activities that we extracted from the server logs.

In addition, the most reliable indicator for the end of an activity was the initiation of a new activity by the user within the same HTTP session. As a consequence, we studied the server log entries and manually adjusted the ends of HTTP sessions when a reasonable duration of non-activity was exceeded. For each participant, in order to minimize times of irrelevant non-activity (e.g., the participant left the computer without logging off), we decided to err on the conservative side by truncating activities and sessions if they appeared exceptionally different from the other patterns of interaction of the respective user.

Finally, the aggregation of 25 into 9 activities was informed by an assessment of the similarity between activities by the authors. In a preliminary analysis of the 25 activities based on Principle Component Analysis (PCA), we found that the boundaries of the components did not coincide with the demarcations that we established between information retrieving, analysis, and deliberation. Instead, PCA suggested the existence of components that varied to the degree in which these three types of activities were intertwined (e.g., various reading activities were grouped with some of the non-deliberative activities).

4.3. Conclusions and future work

In this paper, we discussed two methods for the analysis of human-computer interaction with a web-based GDSS. In particular, we subjected user interaction data of 47 users of the LIT website to multiple sequence alignment and hierarchical cluster analysis. We subscribe to the view that understanding system utilization is a crucial proxy for the assessment of system usefulness which, in turn, is key for the iterative improvement of web-based GDSS. As expressed in the Technology Acceptance Model [9], system utilization hinges on the perceived usefulness of a system in the sense that greater perceived usefulness manifest itself in greater use.

The combination of sequence alignment analysis and hierarchical cluster analysis provides for considerable data analysis synergies. Foremost, once the sequences of individual user interaction are created, activity initiation counts and activity durations can be extracted in a straight-forward manner through the parsing of the text-based sequence files. Such approach offers the researcher two complimentary angles by which behavioral patterns in user interaction data can be explored.

A pragmatic advantage of hierarchical cluster analysis is that hierarchical cluster algorithms are implemented by various statistical software packages which offer functionality for the subsequent analysis of the composition of the clusters. For instance, SPSS automatically appends the cluster membership values to the attribute table of each observation. This is not possible with ClustalTXY as it does not offer data management functionality other than the input and output of sequence files. Instead, one has to manually input cluster membership values into a data file that can be read by other software before further analysis.

Despite having focused on a particular web-based GDSS, we see our work as applicable to any type of web-based system capable of capturing user interaction in a sequential manner. The user interaction data could then be coded as sequences of user-initiated activities and be subjected to sequence alignment analysis. After the extraction of durations of activities from the sequences, hierarchical cluster analysis could be applied to these data. This two-step approach provides a methodological framework for the analysis of system utilization.

In this context, we recommend implementing logging functionality that captures user interaction at the activity level as part of the design of web-based GDSS. For our data, such functionality would have eliminated the need for the conversion of server logs from a list of server-side method invocations to sequences of user activities. In addition, the aspects of user inactivity required that the server logs be screened for sessions that were not terminated by the user. Thus, we suggest incorporating log-off procedures that end a session after a set period of inactivity.

In term of future research, we intend to 1. analyze the clusters from hierarchical cluster analysis regarding the types of user interaction each cluster represents and 2. subject the clusters to logistic regression analysis to test for the existence of statistical associations between user interaction patterns and individual-level characteristics (e.g., age and computer expertise)—the goal of which is to further contribute robust methods for the empirical assessment of web-based GDSS.

- [1] M. B. Ayed, H. Ltifi, C. Kolski, A. M. Alimi, A user-centered approach for the design and implementation of KDD-based DSS: A case study in the healthcare domain, *Decision Support Systems* 50 (1) (2010) 64–78, ISSN 0167-9236.
- [2] E. Mcfadden, D. R. Hager, C. J. Elie, J. M. Blackwell, Remote usability evaluation: Overview and case studies, *International Journal of Human-computer Interaction* 14 (3-4) (2002) 489–502.
- [3] D. Amott, Cognitive biases and decision support systems development: A design science approach, *Information Systems Journal* 16 (1) (2006) 55–78.
- [4] S. Jarupathirun, F. Zahedi, Exploring the influence of perceptual factors in the success of web-based spatial DSS, *Decision Support Systems* 43 (3) (2007) 933–951.
- [5] D. L. Goodhue, R. L. Thompson, Task-technology fit and individual performance, *MIS Quarterly* 19 (2) (1995) 213–236.
- [6] D. L. Goodhue, Understanding user evaluations of information systems, *Management Science* 41 (12) (1995) 1827–1844.
- [7] D. Goodhue, B. D. Klein, S. T. March, User evaluations of IS as surrogates for objective performance, *Information & Management* 38 (2) (2000) 87–101.
- [8] S. Jarupathirun, F. Zahedi, GIS as spatial support systems, in: J. B. Pick (Ed.), *Geographic Information Systems in Business*, Idea Group Publishing, Hershey, PA, 151–174, 2005.
- [9] F. D. Davis, R. P. Bagozzi, P. R. Warshaw, User acceptance of computer technology: A comparison of two theoretical models, *Management Science* 35 (8) (1989) 982–1003.
- [10] F. D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quarterly* 13 (3) (1989) 319–340.
- [11] V. Venkatesh, F. D. Davis, A theoretical extension of the Technology Acceptance Model: Four longitudinal field studies, *Management Science* 46 (2) (2000) 186–204.
- [12] L. A. Le Blanc, K. A. Kozar, An empirical investigation of the relationship between DSS usage and system performance: A case study of a navigation support system, *MIS Quarterly* 14 (3) (1990) 263–277.
- [13] P. Jankowski, T. Nyerges, S. Robischon, K. Ramsey, D. Tuthill, Design considerations and evaluation of a collaborative, spatio-temporal decision support system, *Transactions in GIS* 10 (3) (2006) 335–354.
- [14] C. Rinner, Web-based spatial decision support: Status and research directions, *Journal of Geographic Information and Decision Analysis* 7 (1) (2003) 14–31.
- [15] H. K. Bhargava, D. J. Power, D. Sun, Progress in Web-based decision support technologies, *Decision Support Systems* 43 (2007) 1083–1095.
- [16] T. Zhong, R. K. Young, M. Lowry, G. S. Rutherford, A model for public involvement in transportation improvement programming using participatory Geographic Information Systems, *Computers, Environment and Urban Systems* 32 (2) (2008) 123–133, ISSN 0198-9715.
- [17] M. B. Lowry, T. L. Nyerges, G. S. Rutherford, Internet portal for participation of large groups in transportation programming decisions, *Transportation Research Record* 2077 (2008) 156–165.
- [18] D. Power, S. Kaparthy, Building web-based decision support systems, *Studies in Informatics and Control* 11 (4) (2002) 291–302.
- [19] F. Corpet, Multiple sequence alignment with hierarchical clustering, *Nucleic Acids Research* 16 (22) (1988) 10881–10890.
- [20] D. G. Higgins, P. M. Sharp, CLUSTAL: a package for performing multiple sequence alignment on a microcomputer, *Gene* 73 (1) (1988) 237–244.
- [21] E. Zuckerkandl, L. Pauling, Molecules as documents of evolutionary history, *Journal of Theoretical Biology* 8 (2) (1965) 357–366.
- [22] R. F. Doolittle, Similar amino acid sequences: Chance or common ancestry?, *Science* 214 (4517) (1981) 149–159.
- [23] L. Wang, T. Jiang, On the complexity of multiple sequence alignment, *Journal of Computational Biology* 1 (4) (1994) 337–348.
- [24] W. Just, Computational complexity of multiple sequence alignment with SP-score, *Journal of Computational Biology* 8 (6) (2001) 615–623.
- [25] I. Elias, Settling the intractability of multiple alignment, *Journal of Computational Biology* 13 (7) (2006) 1323–1339.
- [26] A. Abbott, Sequence analysis: New methods for old ideas, *Annual Review of Sociology* 21 (1995) 93–113.
- [27] A. Abbott, A primer on sequence methods, *Organization Science* 1 (4) (1990) 375–392.
- [28] A. Abbott, A. Hrycak, Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers, *American Journal of Sociology* 96 (1) (1990) 144–185.
- [29] C. Wilson, A. Harvey, J. Thompson, ClustalG: Software for analysis of activities and sequential events, in: Paper presented at the Workshop on Longitudinal Research in Social Science: A Canadian Focus, London, Ontario, Canada, London, Ontario, Canada, 1999.
- [30] C. Wilson, Activity patterns in space and time: Calculating representative Hagerstrand trajectories, *Transportation* 35 (2008) 485499.
- [31] D. G. Higgins, A. J. Bleasby, R. Fuchs, CLUSTAL V: Improved software for multiple sequence alignment, *Computer Applications in the Biosciences* 8 (2) (1992) 189–191.
- [32] J. D. Thompson, D. G. Higgins, T. J. Gibson, CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through

- sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Research* 22 (22) (1994) 4673–4680.
- [33] D. G. Higgins, J. D. Thompson, T. J. Gibson, Using CLUSTAL for multiple sequence alignments, *Methods in Enzymology* 266 (1996) 383–402.
- [34] R. Chenna, H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, J. D. Thompson, Multiple sequence alignment with the Clustal series of programs, *Nucleic Acids Research* 31 (13) (2003) 3497–3500.
- [35] R. C. Edgar, S. Batzoglou, Multiple sequence alignment, *Current Opinion in Structural Biology* 16 (3) (2006) 368–373.
- [36] C. Wilson, Analysis of travel behavior using sequence alignment methods, *Transportation Research Record* 1645 (1998) 52–59.
- [37] C. Wilson, Activity patterns of Canadian women: Application of ClustalG sequence alignment software, *Transportation Research Record* 1777 (1) (2001) 55–67.
- [38] B. Bargeman, C.-H. Joh, H. Timmermans, Vacation behavior using a sequence alignment method, *Annals of Tourism Research* 29 (2) (2002) 320–337.
- [39] N. Shoval, M. Isaacson, Sequence alignment as a method for human activity analysis in space and time, *Annals of the Association of American Geographers* 97 (2) (2007) 282–297.
- [40] S. I. Fabrikant, S. Rebich-Hespanha, N. Andrienko, G. Andrienko, D. R. Montello, Novel method to measure inference affordance in static small-multiple map displays representing dynamic processes, *The Cartographic Journal* 45 (3) (2008) 201–215.
- [41] M. Chen, Y. Liou, C.-W. Wang, Y.-W. Fan, Y.-P. J. Chi, TeamSpirit: Design, implementation, and evaluation of a Web-based group decision support system, *Decision Support Systems* 43 (4) (2007) 1186–1202.
- [42] R. Aguirre, T. Nyerges, Geovisual evaluation of public participation in decision making: The grapevine, *Journal of Visual Languages & Computing* 22 (4) (2011) 305–321.
- [43] T. Nyerges, R. W. Aguirre, Public participation in analytic-deliberative decision making: Evaluating a large-group Online field experiment, *Annals of the Association of American Geographers* 101 (3) (2011) 561–586.
- [44] R. Young, T. Zhong, M. Lowry, G. S. Rutherford, T. Nyerges, An analytical-deliberative online framework for transportation programming, *International Journal of Technology, Knowledge, and Society* 3 (2) (2007) 89–98.
- [45] R. Johnson, D. Wichern, *Applied multivariate statistical analysis*, Pearson Prentice Hall, New Jersey, NJ, 2007.
- [46] J. H. Ward, Hierarchical grouping to optimize an objective function, *Journal of the American Statistical Association* 58 (301) (1963) 236–244.
- [47] R. Mojena, Hierarchical grouping methods and stopping rules: An evaluation, *The Computer Journal* 20 (4) (1977) 359–363.
- [48] A. El-Hamdouchi, P. Willett, Comparison of hierarchic agglomerative clustering methods for document retrieval, *The Computer Journal* 32 (3) (1989) 220–227.
- [49] J. Felsenstein, Confidence limits on phylogenies: An approach using the bootstrap, *Evolution* 39 (4) (1985) 783–791, ISSN 00143820.
- [50] J. D. Thompson, F. Plewniak, O. Poch, A comprehensive comparison of multiple sequence alignment programs, *Nucleic Acids Research* 27 (13) (1999) 2682–2690.

5. Acknowledgments

This research was supported by the National Science Foundation Information Technology Research Program Grant No. EIA 0325916 which we gratefully acknowledge. We would also like to acknowledge the PGIST research team and especially Tim Nyerges, Matt Wilson, Kevin Ramsey, Robert Aguirre, Michael Lowry, Arika Lingmann-Zielinska, Michael Patrick, Zhong Wang, Jie Wu, and Guirong Zhou for their contributions to the project. Lastly, we thank the editor and the reviewers for their support of and thorough comments on our manuscript. The authors are solely responsible for the content.

List of figure captions

Figure 1: An example of a server log entry (excluding non-crucial parameters).

Figure 2: An example of a user interaction sequence file in the FASTA format for two users with the identifier 13 and 26. The sequence of user 13 is comprised of 54 and the sequence of user 26 of 41 three-character words.

Figure 3: An example of an icicle plot for the furthest neighbor clustering method. Unique user identifiers are shown along the x-axis and numbers of clusters along the y-axis. For instance, read from left to right, a two-cluster solution is comprised of user 47, 46, 42, and 44 in the first and the remaining users in the second cluster.

Figure 4: Bootstrap results for the activity successions (i.e., no duration) of 25 and 9 activities. The leaves of the tree denote unique case (i.e., user) IDs. The lengths of the tree branches convey the standardized distance (i.e., dissimilarity) between cases, with values ranging from 0 to 0.5, in increments of 0.1.

Figure 4 (a): 25 activities

Figure 4 (b): 9 activities

Figure 5: Bootstrap results for the activity durations of 25 and 9 activities. The leaves of the tree denote unique case (i.e., user) IDs. The lengths of the tree branches convey the standardized distance (i.e., dissimilarity) between cases, with values ranging from 0 to 0.5, in increments of 0.1.

Figure 5 (a): 25 activities

Figure 5 (b): 9 activities

List of table captions

Table 1: Overview of the LIT steps with their respective availability dates.

Table 2: User interaction activities overview for the fine (i.e., 25 activities) and aggregated (i.e., 9 activities) levels of analysis.

Table 3: User interaction activities coding schema at the level of 25 and 9 activities.

Table 3 (a): fine-level user interaction

Table 3 (b): aggregated user interaction

Table 4: Overview of hierarchical cluster (HCA) and multiple sequence alignment analyses (MSA) of the LIT data in relation to the granularity of activities. Individual analyses are denoted with an x in the cells. Appropriate pairwise comparisons of the classification results from HCA and MSA are represented through arrows linking the respective activity analysis levels.

Table 5: User activity analyses overview showing the four possible combinations of the sequence alignment analysis in relation to the matched comparative hierarchical cluster analyses. The variable under analysis, the distance measure, and the clustering algorithms that were employed in the hierarchical cluster analyses are stated in the cells of the table.

Table 6: Hierarchical cluster analysis results for the three activity levels in relation to varying clustering methods. A white square indicates a weak classification (i.e., clusters of unbalanced size, including a number of single-case clusters) and a black square a strong classification (i.e., well-formed clusters of relatively comparable size). A black triangle represents a classification that is considered to be in-between weak and strong. A strong classification is desirable for a statistical analysis of the characteristics of the cases in the clusters.

Biographical note:

<p>Dr. Martin Swobodzinski</p> <p>Assistant Professor Director, Center for Spatial Analysis and Research Portland State University Department of Geography Portland, OR 97201</p> <p>Office: 424I Cramer Hall Phone: (503) 725-3164 e-mail: swobod@pdx.edu</p> <p><u>Education</u> Ph.D., UC-Santa Barbara and San Diego State University, 2012 Dipl., Geoinformatics, University of Muenster, Germany, 2006</p> <p><u>Research interests</u></p> <ul style="list-style-type: none"> ▪ Geographic information Science ▪ Behavioral Geography ▪ Spatial Decision Support ▪ Human-Computer Interaction ▪ Transportation 	<p>Dr. Piotr Jankowski</p> <p>Professor, Department Chair San Diego State University Department of Geography San Diego, CA 92182-4493</p> <p>Office: Geography Annex 208 Phone: (619) 594-0640 e-mail: pjankows@mail.sdsu.edu</p> <p><u>Education</u> Ph.D., University of Washington, 1989 M.S., University of Economics, Poznan, Poland, 1979</p> <p><u>Research interests</u></p> <ul style="list-style-type: none"> ▪ Spatial Decision Support Systems ▪ Public Participation GIS ▪ Visual Analytics ▪ Geocomputation Methods ▪ Volunteered Geographic Information
--	---

Highlights:

- System utilization is crucial for the assessment of the performance of DSS.
- An exploratory analysis of user interaction with web-based GDSS is presented.
- Clusters of user interaction from two clustering methods are derived and compared.
- Multiple sequence alignment analysis resulted in non-reliable clusters.
- Hierarchical cluster analysis proved reliable clusters for overall interaction duration.